

REMARKS

Applicant notes the indication of claims 5 and 10 containing allowable subject matter.

Claims 9 and 11 have been cancelled to expedite prosecution.

Claims 1 and 8 have been amended for clarity.

Applicant traverses the rejection of claims 1-4, 6-8 as being unpatentable under 35 USC 103(a) as being obvious as a result of Van den Akker (USP 6,415,250) in view of de Campos (USP 6,272,456). Paragraph [0014] of applicant's published application indicates the language identification system of Van den Akker (US Patent 6,415,250) is limited to a single category of first character strings, such as suffixes (prefixes), in a word. Therefore, Van den Akker does not include the requirement of independent claims 1 and 8 to analyze words extracted from digital text, to thereby construct for each extracted word plural character strings contained in said extracted word, including prefixes, suffixes and infixes, with overlap and different lengths lying between one character and the number of characters in said extracted word. The Van den Akker analyzer analyzes only one character string per extracted word relative to the corpus of a language.

The significance of the foregoing difference between applicant's claims 1 and 8 and Van den Akker can be seen by comparing certain parts of the Van den Akker specification and applicant's specification.

In Van den Akker [column 11, line 65 - column 12, line 18], a "suffix of the each of the parsed words 403" in the suffix extractor 404 "is determined to have a predetermined number of characters at the end of each parsed word," and "the predetermined number of characters in the parsed words 403 extracted by the suffix extractor 404 is three characters," and "the source language of an unknown text 301" is identified "by analyzing the last three characters of the parsed words 403," and "for example, 4 characters at the end of the word may be used to capture the suffix." In the same manner, the predetermined number of characters, three or four, defines the length of a prefix in the parsed word [column 8, line 58 to column 9, line 3; "word portions that contains the prefix may used"; Fig. 2C].

In the applicant's specification [0032 of applicant's published application], "the first three directors PRq, SUq and INq relate to morphemes, syllables and short character strings CH of from one to six characters, for example" and more particularly relate respectively to prefixes, suffixes and infixes in the predetermined languages. An infix is defined as a character string that is included between the first character, i.e., the start, of a word, and the last character, i.e., the end, of the word, as e.g., "ou" or "oi" [0036 of applicant's published application]. In other examples [Van den Akker, column 8, lines 53-57; see also Figs. 2B-2C], the applicant's directory SUq includes the suffix "ing" of the word "smashing," but also infixes "mash" and "ash" of the "smashing"; and the applicant's directory SUq includes the suffix "ment" of the word "development," but also infixes "velo," "ve," and "lo" of the word "development."

As a result of the above remarks, the word analyzing means in the claimed device constructs a plurality of character strings contained in an extracted word and having lengths lying between one character and the number of characters in said extracted word, such as the prefix "de," the suffix "ment" and other character chains "eve," "evelo," "elo," "lopme," "opme," etc., including infixes "velo," "ve," and "lo," included in the extracted word "development" and thus including chains partially overlapping ([0057] of applicant's published application: "In the latter variant, the character strings CH contained in the extracted word MT and found in the directories PRq, SUq and INq may partially overlap. This is in contrast to the n-grams of the approach disclosed by US Patent No. 6,292,772 B1 already commented on. For example, if the processed word MT is the French word "aiment," the character strings "ment" and "ent" placed in the pseudo-suffix director SUq overlap in the processed word. To cite another example, the infix "oi" and the pseudo-suffix "is" of the processed word "vois" overlap.")

Therefore, applicant again states that Van den Akker fails to show the function of the claimed analyzing means and more particularly the means for analyzing words extracted from said digital text thereby constructing for each extracted word a plurality of character strings, including prefixes (claim 2: including pseudo-prefixes), suffixes (claim 2: including pseudo-suffixes) and infixes (character chains included between the start and end of said extracted word) with overlap (e.g. infixes and pseudo-infixes) and

different lengths lying between one character and the number of characters in said extracted word.

Van den Akker also fails to disclose the requirements of claims 1 and 8 for (1) prestorage first character strings, including prefixes, suffixes and infixes of different lengths from words of a plurality of predetermined languages, as character chains included between the start and end of said words and overlapping (including infixes and pseudo-infix infixes), that occur frequently anywhere respectively in said words of said predetermined languages, and (2) prestorage second character strings of different lengths (including prefixes and suffixes in said words of said plurality of predetermined languages and character chains included between the start and end of said words and overlapping (e.g. infixes and pseudo-infix infixes)), that are atypical anywhere respectively in said words of said predetermined languages.

Applicant also disagrees with the contention in the office action that Van den Akker bases the score of a character chain extracted from a parsed word based on the position of the first character chain in the extracted of a parsed word. For example, said first character chain of a Van den Akker parsed word is a suffix (or a prefix) and always has the same position in the extracted words, i.e., the last position in all the parsed words, ending with the suffix, counted from the last character in the parsed words (or for a prefix, the first position in all the parsed words starting by said prefix, counted from the first character in the parsed words).

The position of a character chain in a parsed word is not disclosed by Van den Akker, in the paragraph in column 9, lines 18-41 cited by the examiner. Van den Akker only relates to the frequency of a character chain in words of a language corpus. A selected word portion (character chain), such as a suffix, also called a "word ending" (or a prefix) "was extracted from the corresponding language corpus 309 along with a frequency value indicative of the number of times the selected word portion was found within the corresponding language corpus 309. These frequencies are subsequently normalized to a common corpus size, serving as relative frequencies of each of the selected word portion in each language" [column 9, lines 34-41]; also see the frequency lists in column 13, lines 33-42.

Applicant's claims 1 and 8 indicate the position and frequency of a character chain in an extracted word are distinctly identified by the applicant, as disclosed by (0059) of applicant's published application. For example, the coefficients PO, FR and LON respectively depend on the position of the character string CH in the extracted word MT, on the frequency of the character string CH in the determined language Lq, and on the length of the character string CH. For example, in the extracted words "development," "abutment" and "government," the suffix "ment" always has the last position. In contrast, Van den Akker does not consider any distinction relating to the position of the character chain "ment" in these three words, while the applicant assigns different first coefficients depending on the positions 8, 4, 3, 6 and 5 of the character chain "ment," counted from the beginning of the four words ""development," "alimented," "lamentable," "incremental" and "rudimental," or the positions 4, 6, 8, 6 and 6 of the character chain "ment" counted from the ends of the five words respectively. In a corpus including the words "development," "abutment," "government," "alimented," "lamentable," "incremental" and "rudimental," the frequency of the character chain "ment" is 7 and different of the values of the positions.

Therefore, Van den Akker fails to the requirement of claims 1 and 8 for: whenever a first character string is found in said extracted word, a score associated with one determined language increased by a first coefficient depending on the position of said first character string found in said extracted word.

Van den Akker [column 13, line 62 to column 14, line 24] effectively discloses that the probability, i.e., the normalized frequency value, of a word portion can be negative to indicate a strong likelihood that the language of the input text is not the corresponding language. But the Van den Akker word portion is a suffix, while a second coefficient in claims 1 and 8 is associated with a second character string which can be a prefix, a suffix or any word chain included between the start and end of words of the predetermined languages and overlapping (including infixes and pseudo-infix infixes) [0042-0044 of applicant's published application].

Because of the foregoing differences between claims 1 and 8 and Van den Akker that are not considered in the office action, there is no need to discuss de Campos that the examiner alleges concerns the second coefficient associated with a second

character string that depends on a predetermined language. This is particularly the case because neither Van den Akker nor de Campos discloses the requirements of claims 1 and 8 relating to a first coefficient that depends on the position of a character string found in an extracted word.

Because claims 1 and 8 are allowable over Van den Akker in view of de Campos dependent claims 2-4, 6 and 7 are also allowable.

To the extent necessary, a petition for an extension of time under 37 C.F.R. 1.136 is hereby made. Please charge any shortage in fees due in connection with the filing of this paper, including extension of time fees, to Deposit Account 07-1337 and please credit any excess fees to such deposit account.

Respectfully submitted,

LOWE HAUPTMAN HAM & BERNER, LLP

/Allan M. Lowe/

Allan M. Lowe
Registration No. 19,641

USPTO Customer No. 22429
1700 Diagonal Road, Suite 300
Alexandria, VA 22314
(703) 684-1111
(703) 518-5499 Facsimile
Date: March 18, 2009
AML/cjf